Author: Veronika Andrushchenko

Final Project Report: Methodological insights and sentiment analysis of Lithuanian news on school examinations

1. Introduction

Education tends to be a catalyst for a country's economic development, as it prepares a stable, skilled, and knowledgeable generation of young people who later form the foundation of the labour force. Because of this, the field is continuously evolving through reforms, new policies, and changes to the examination system, all of which are widely represented in the media. The latter plays a crucial role in shaping public perceptions of these dynamics and developments, making it essential to analyse the emotional tone of news coverage – a task that can be approached through sentiment analysis (SA). However, most existing sentiment models are developed for English, leaving low-resource languages such as Lithuanian underexplored.

The goal of this project is to advance methodological understanding of sentiment analysis in media studies by integrating bibliometric mapping, comparative evaluation of analytical approaches, and empirical application to Lithuanian news coverage of school examinations. To achieve this goal, the project is guided by the following **objectives**:

1. To conduct a bibliometric analysis of scholarly research on the representation of education in media, with the aim of systematically identifying key authors, influential works, thematic clusters, and emerging research trends;
2. To critically assess sentiment analysis methodologies by employing Qualitative Comparative Analysis (QCA) as a systematic approach for evaluating their accuracy, scalability, and applicability in non-English language contexts;
3. To apply sentiment analysis methodologies to a corpus of Delfi.lt news articles covering school examinations in Lithuania between 2020 and 2025, to empirically evaluate prevailing emotional tones and assess methodological performance in a real-world context.

The main **research questions** that this study aims to answer are:

1. What is the overall sentiment distribution of Lithuanian news articles on school examinations?
2. How does sentiment change over time across 2020–2025?
3. What lexical or thematic differences characterize positive vs. negative articles?
4. How suitable is the lexicon-based method for analysing Lithuanian news?

The findings aim to provide the sentiment-based overview of Lithuanian examination-related news, demonstrate methodological feasibility, challenges and limitations for low-resource languages, and offer recommendations for further improvement.

2. Background

Understanding emotional patterns in educational news requires placing the empirical analysis within the broader scope of developments in the field. While this report does not present the full bibliometric and methodological investigation conducted in the larger project, the most relevant insights needed to provide background, namely a bibliometric analysis of media and education publications and Qualitative Comparative Analysis of sentiment analysis methodologies, are summarised here.

2.1 Bibliometric Landscape of Education and Media Research (2015–2025)

A bibliometric analysis conducted in Scopus initially identified over 200,000 publications related to education and media, which after applying filters for years (2015–2025), language (English), document type (articles, reviews, conference papers), and disciplinary alignment (social sciences, humanities) was reduced to 28,558 records. Subsequent cleaning to remove missing data, duplicates, and outliers resulted in a final dataset of 1,360 publications suitable for bibliometric analysis.

The analysis used the following applications: Bibliometrix, an R-based package designed for quantitative science mapping and bibliometric statistics, and VOSviewer, a software tool used for constructing and visualising bibliometric networks.

Starting from the main trend in education and media, the following shifts were identified:

- Methodological shift toward computational approaches, including sentiment analysis, natural language processing, machine learning, and topic modelling;
- Growth in crisis-related educational topics, including COVID-19 disruptions, e-learning, and the transformation of educational assessments;
- Use of SA in studies of public opinion, media framing, and education reforms (Figure 1).

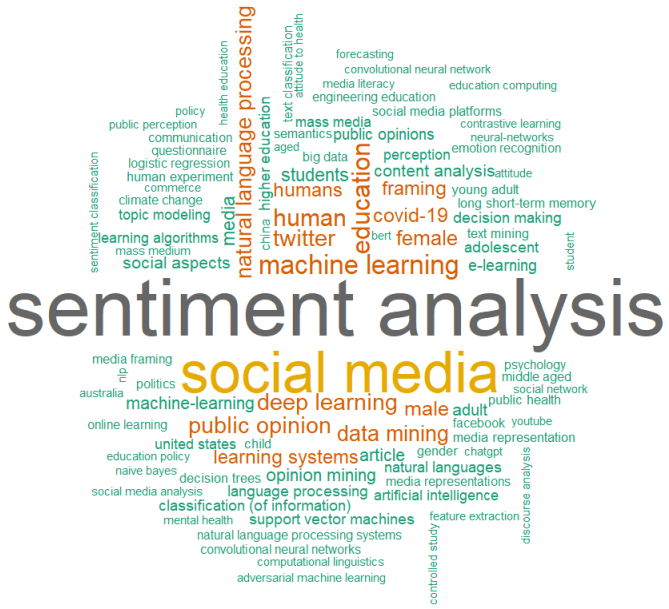

Figure 1. Word cloud of the most frequent author keywords (Bibliometrix).
Source: Author's bibliometric analysis based on Scopus records.

2.2 Key Authors, Influential Works, and Collaboration Networks

As about key authors, the field is driven by a relatively small number of scholars, who can be divided into the following groups:

- Leading authors with high productivity and strong citation impact (e.g., Grljević, Kovačević, Cambria);
- Consistent contributors with steady output and moderate impact (e.g., Ong, Ihediwa);
- Emerging authors with recent but increasing activity (e.g., Budiwati, Nguyen) (Figure 2, Table 1).

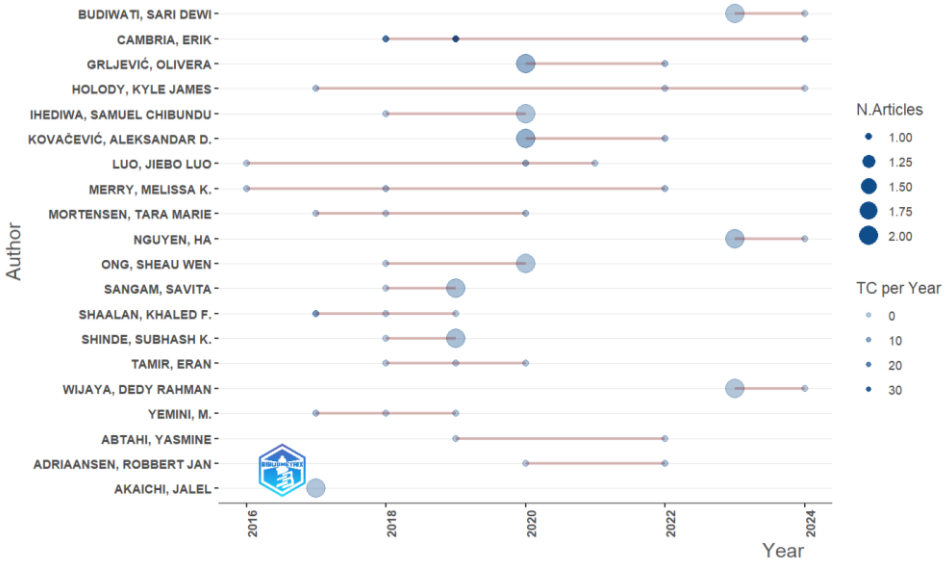

Figure 2. Authors' production over time, 2016–2024.

Source: Author's bibliometric analysis (Bibliometrix).

Table 1. Top authors by yearly productivity and citation performance.

Author	Year	Publications	Total Citations	Citations per Year
GRLJEVIC, OLIVERA	2020	2	45	7,5
KOVACEVIC, ALEKSANDAR D.	2020	2	45	7,5
SANGAM, SAVITA	2019	2	21	3
SHINDE, SUBHASH K.	2019	2	21	3
AKAICHI, JALEL	2017	2	10	1,1
IHEDIWA, SAMUEL CHIBUNDU	2020	2	10	1,7
NGUYEN, HA	2023	2	10	3,3
ONG, SHEAU WEN	2020	2	10	1,7
BUDIWATI, SARI DEWI	2023	2	3	1
WIJAYA, DEDY RAHMAN	2023	2	3	1

Source: Scopus dataset processed with Bibliometrix (author's analysis).

Co-authorship networks, in turn, revealed tightly connected clusters, indicating strong cross-disciplinary collaboration structured around a few central figures (such as Carpenter and Greenhow), namely:

- Digital tools and professional development cluster (Carpenter, Trust, Krutka);
- Social media and online learning cluster (Greenhow, Daly, Galvin);
- Media literacy and networked education cluster (Manca, Ranieri);
- Innovation and digital pedagogy cluster (Marín) (Figure 3).

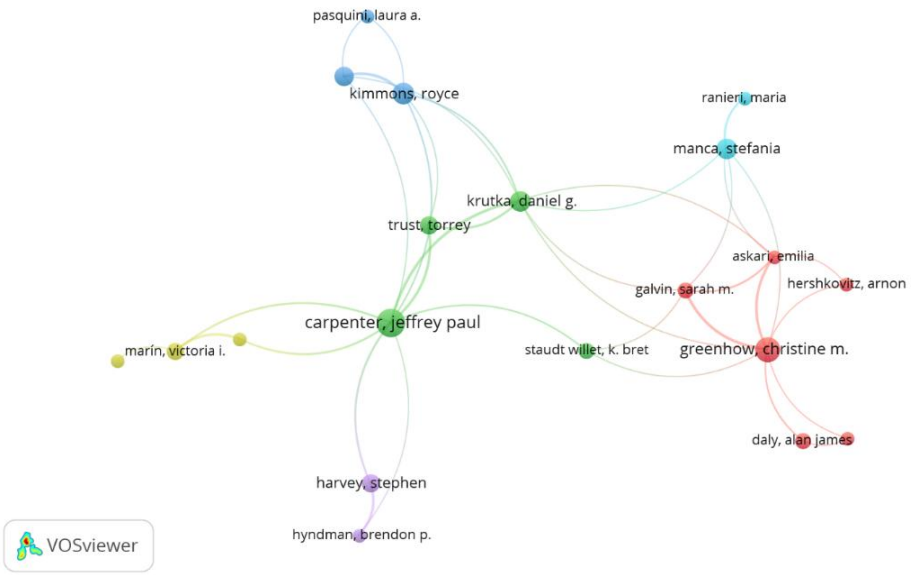

Figure 3. Co-authorship network of authors (VOSviewer, threshold ≥ 5 documents).
Source: Author's VOSviewer analysis.

At the same time, research on education and media is globally distributed but driven by a limited number of influential contributors, which can be confirmed by the following research hubs:

- North America (Michigan State University, Stanford University);
- Asia (University of Hong Kong, Universitas Pendidikan Indonesia);
- Europe (KU Leuven, Complutense University of Madrid, Kazan Federal University);

- Australia–UK partnerships (Deakin University, Monash University, University of Oxford) (Figure 4, 5).

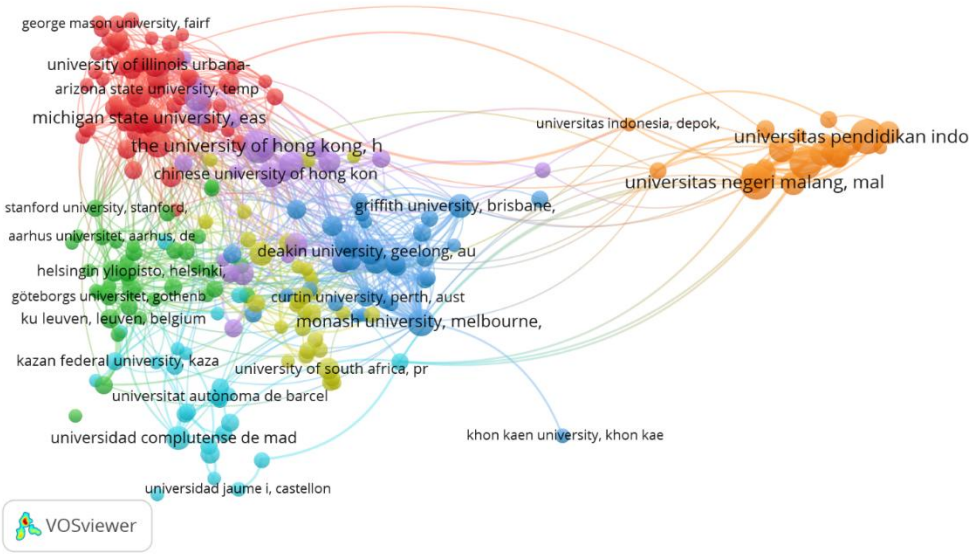

Figure 4. Co-authorship network of organizations (VOSviewer, threshold ≥ 20 documents). Source: Author's VOSviewer analysis.

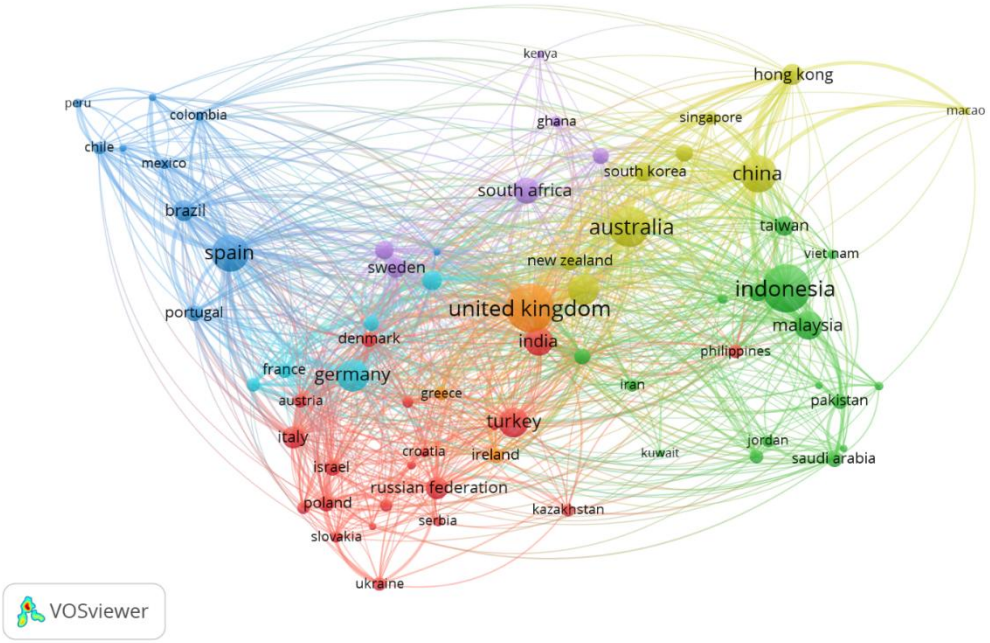

Figure 5. Country-level co-authorship network; VOSviewer, threshold ≥ 35 documents. Source: Author's VOSviewer analysis.

2.3 Thematic Structure and Keyword Clusters

Keyword co-occurrence analysis revealed the following thematic clusters, with media literacy serving as a bridge concept**:**

- Student experience, identity, and social contexts;

- Technology-supported learning (AR/VR, computational thinking, innovation);
- Teacher education and educational policy;
- Data-driven and analytics perspectives (learning analytics, risk perception);
- Media-rich environments (digital storytelling, social networks, cyberbullying) (Figure 6).

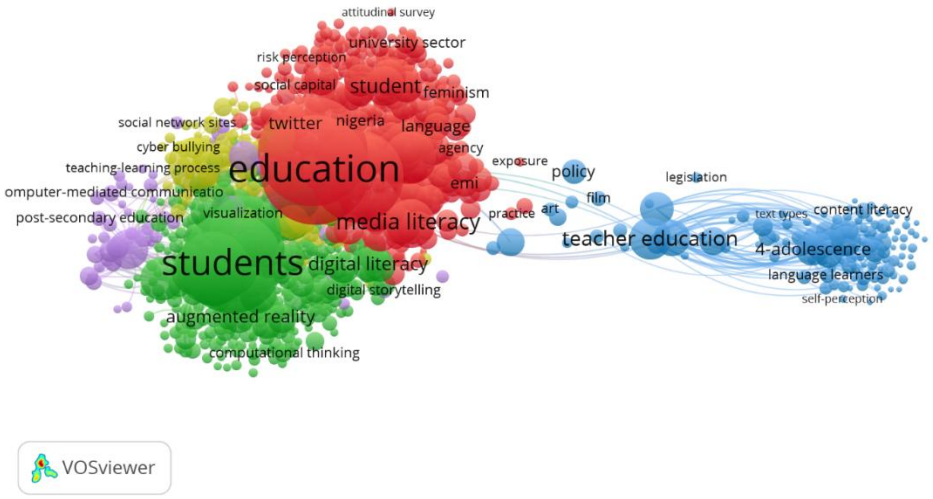

Figure 6. Keyword co-occurrence network and clusters (VOSviewer, threshold ≥ 5 occurrences).

Source: Author's VOSviewer analysis.

At the same time, word frequency analysis revealed the field's shift toward data-intensive research, with the following keywords being the most common: sentiment analysis (581), social media (392), machine learning (140), deep learning (111) (Table 2).

Table 2. Top author keywords by frequency.

Keyword	Frequency
sentiment analysis	581
social media	392
social networking (online)	210
machine learning	140
education	128
human	116
deep learning	111
twitter	110
public opinion	108
data mining	97

Source: Scopus dataset processed with Bibliometrix (author's analysis).

2.4 Emerging Research Trends

Finally, overlay maps illustrated a clear chronological progression in the given field:

- Pre-2018: focus on traditional topics (teacher identity, media literacy, communication);
- 2019–2020: rapid rise of machine learning, NLP, sentiment analysis;
- 2020 onward: surge in crisis-related research (COVID-19) and examination of remote learning;
- 2022 onward: growth of hybrid analytical approaches, digital storytelling, and multilingual education studies (Figure 7).

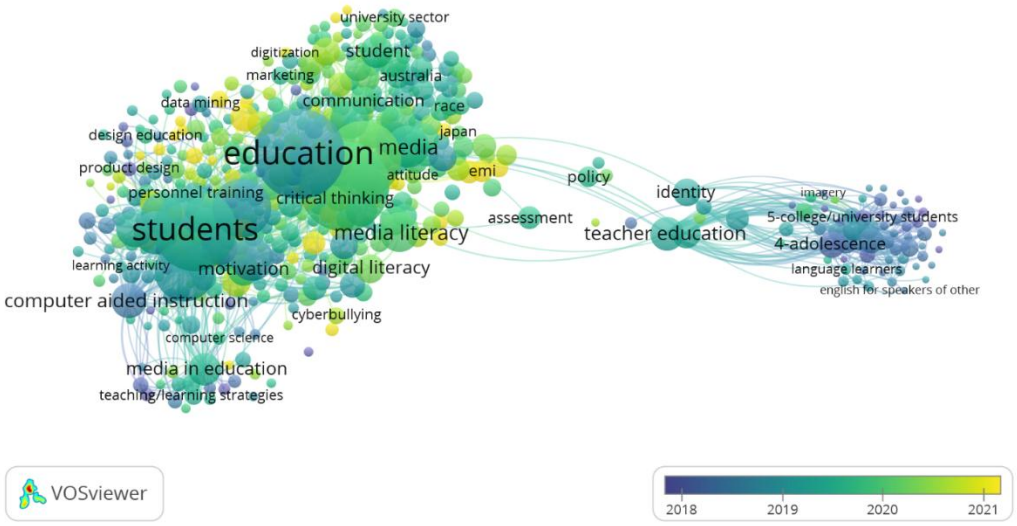

Figure 7. Overlay map by average publication year; VOSviewer.
Source: Author's VOSviewer analysis.

Overall, the bibliometric analysis revealed a shift of media and education studies from traditional qualitative approaches toward computational methods, with growing reliance on data-driven methodologies, large-scale textual datasets, and automated techniques such as sentiment analysis, topic modelling, and machine learning.

2.5 Qualitative Comparative Analysis (QCA) of Sentiment Analysis Approaches

Together with the bibliometric analysis, a comparative assessment of SA approaches, namely lexicon-based methods, machine learning models, deep learning (transformers), and hybrid techniques, was conducted to evaluate their suitability for analysing Lithuanian educational news. Based on classical and recent surveys listed in the References, a QCA was applied using five conditions: accuracy, scalability, adaptability to non-English languages, data requirements, and interpretability (Table 3).

Table 3. Conceptual QCA Calibration Matrix.

Method	Accuracy	Scalability	Adaptability	Data Need	Interpretability
Lexicon-based	×	✓	△	✓	✓
Machine Learning	✓	✓	×	×	△

Transformers	✓	×	✓	×	×
Hybrid	△	△	△	△	×

Source: Author's elaboration based on references [1-7].

The QCA results revealed that lexicon-based methods tend to be the most favourable for low-context languages, with machine learning or transformers as the second option, and hybrid as the least applicable. Considering simplicity, transparency, scalability and limited availability of Lithuanian linguistic resources, the lexicon-based method was selected as the most practical option for further application.

3. Methods

The analysis begins with collecting and preparing a dataset of Lithuanian news articles about national school examinations. The workflow includes data scraping, cleaning, filtering, and applying a lexicon-based sentiment analysis model, with each step being described below.

3.1 Dataset Description and Collection

The dataset used for the sentiment analysis consists of news articles published on Delfi.lt, one of Lithuania's largest media portals. Articles were collected with the Web Scraper browser extension, which allows automated extraction of manually selected page elements, based on the following keyword groups:

- "valstybini brandos egzamina"** (state maturity examinations);
- "pagrindinio ugdymo pasiekimų patikrinim"* (PUPP – lower secondary achievement testing);
- "nacionalini mokinių pasiekimų patikrinim"** (national student assessment).

In this project, each news article is treated as a **node**, while the **attributes** are: publication date (and extracted year), article text, sentiment scores (neg, neu, pos, compound), sentiment label (positive, neutral, negative), URL (unique identifier).

The initial corpus accounted for 272 articles before cleaning (Table 4).

Table 4. Example rows from the cleaned Delfi.lt examination-related news dataset.

title	date	text	url	year
PrasidД—jus karantinui iEЎkilo klausimas dД—l abiturie...	*07.04.2020 12:32:00*	вЂћJaunas Esmogus, baigdamas mokyklД..., gauna bran...	https://www.delfi.lt/projektai/studijos/prasidejus-kara...	*2020*
MokiniEi pasiekimEi reitingai: prestiEsiniEi gimnazijEi tr...	*06.05.2020 11:01:00*	Geriausiomis EЎalies gimnazijomis, kaip ir pernai, pripa...	https://www.delfi.lt/news/daily/lithuania/mokiniu-pasi...	*2020*
Lietuva jau dabar ruoEЎiasi blogiausiam scenarijui: bran...	*07.09.2020 14:10:00*	PokyДЌiai patikrinimuose ir egzaminuose Pagrindinio u...	https://www.delfi.lt/news/daily/education/lietuva-jau-...	*2020*
Penktadaliu iEЎaugД™s EЎvietimo finansavimas nepage...	*15.09.2020 12:41:00*	Anot iEЎvados, 2016вЂ“2019 metais mokymui skiriamo...	https://www.delfi.lt/news/daily/lithuania/penktadaliu-i...	*2020*
Koalicijos derybos ДЇsibД—gД—ja: skelbiamas visas Lai...	*29.10.2020 11:41:00*	вЂћLaisvД—s partija yra pasiruoEЎusi dalyvauti centro...	https://www.delfi.lt/news/daily/lithuania/koalicijos-de...	*2020*
Saulius JurkeviДЌius: ar reikalingi EЎiais metais brandos...	*02.04.2021 10:52:00*	Daugelis turbE«t nustebs suEsinojД™, kad Lietuvoje bra...	https://www.delfi.lt/news/daily/lithuania/saulius-jurke...	*2021*
E MSM: siE«loma maEsinti brandos egzaminEi reikEЎm...	*20.04.2021 16:22:00*	SiE«lomEi pokyДЌiEi tikslas вЂ“ uEstikrinti sД...lygas bai...	https://www.delfi.lt/news/daily/lithuania/smsm-siulo...	*2021*
Albinas Bagdonas. Brandos egzaminai 2021: matematik...	*27.04.2021 15:12:00*	Sutinku, kad ryEЎys tarp mД...stymo ir matematiniEi ge...	https://www.delfi.lt/news/ringas/lit/albinas-bagdonas-...	*2021*
Prasideda 4 ir 8 klasiEi mokiniEi matematikos ir skaitym...	*03.05.2021 12:17:00*	Ji skelbia parengusi elektroninio nacionalinio mokiniEi p...	https://www.delfi.lt/news/daily/lithuania/prasideda-4-i...	*2021*
Naujausiuose mokyklEi reitinguose вЂ“ privaДЌiEi gimn...	*12.05.2021 10:00:00*	вЂћReitingEiвЂњ Esurnalo vyr. redaktorius Gintaras Sar...	https://www.delfi.lt/news/daily/education/naujausiuos...	*2021*
GimnazijEi reitingas atsakД— ДЇ klausimД..., ar mokym...	*12.05.2021 11:47:00*	Kaip raEЎoma treДЌiadienДЇ paskelbtame praneEЎime,...	https://www.delfi.lt/news/daily/lithuania/gimnaziju-rei...	*2021*
PrivaДЌiam universitetui pamatus klojД™s Chmieliausk...	*09.10.2021 18:47:00*		https://www.delfi.lt/verslo-poziuris/specialus-interviu/...	*2021*
MokiniEi laukia gausybД— pokyДЌiEi: planuoja keisti br...	*03.11.2021 11:17:00*	Vienas iEЎ pagrindiniEi ministerijos planEi viduriniame...	https://www.delfi.lt/news/daily/lithuania/mokiniu-lauk...	*2021*
E iugEsdinienД—: vaikams privalome sukurti sД...lygas...	*14.12.2021 17:00:00*	E MSM teigimu, tai uEstikrintEi E vietimo, mokslo ir spo...	https://www.delfi.lt/news/daily/lithuania/siugzdiniene-...	*2021*
E MSM: brandos atestatui вЂ“ trys privalomi egzaminai...	*14.12.2021 18:18:00*	Egzaminai sudarytEi 60 proc. ДЇvertinimo Kaip antradie...	https://www.delfi.lt/news/daily/lithuania/smsm-brand...	*2021*
Jurgita E iugEsdinienД—. KД... EЎvietime pakeis per EЎi...	*16.12.2021 16:42:00*	NД— vienas vaikas negali bE«ti pamirEЎtas E ios Vyriau...	https://www.delfi.lt/news/ringas/politics/jurgita-siugzd...	*2021*
Jurgita E iugEsdinienД—. KodД—l sprendimai dД—l mo...	*11.05.2022 14:36:00*	Kiek politikEi вЂ“ Seimo nariEi, savivaldybiEi tarybEi nar...	https://www.delfi.lt/news/ringas/politics/jurgita-siugzd...	*2022*
Tyrimo rezultatai keiДЌia darbo rinkos Esaidimo taisykl...	*12.05.2022 6:01:00*	DarbdaviEi lE«kestis вЂ“ praktiEЎkumas E vietimo eksp...	https://www.delfi.lt/projektai/studijos/tyrimo-rezultat...	*2022*
Matematikos egzaminEi rezultatai: вЂћNesiД—mus per...	*12.07.2022 10:22:00*	Vilniaus universiteto (VU) Matematikos ir informatikos...	https://www.delfi.lt/projektai/studijos/matematikos-e...	*2022*
E iugEsdinienД—: atsisveikiname su tuo, kas buvo tapД...	*29.08.2022 10:27:00*	вЂћTai iEЎskirtiniai metai mE«sEi mokyklose. Po maEsd...	https://www.delfi.lt/news/daily/lithuania/siugzdiniene-...	*2022*

Source: Author's own dataset, scraped from Delfi.lt using Web Scraper and processed in Python.

3.2 Dataset Loading and Cleaning

After the scraping stage, the three examination-related datasets (delfi_nacionaliniai.xlsx, delfi_pagrindinis.xlsx, and delfi_valstybiniai.xlsx) were imported into Python. Because the Web Scraper extension stores the article URL in a non-standard column (data-page-selector), each file was first harmonised by renaming this field to url. Only the relevant variables (title, date, text, url) were retained to ensure a uniform structure across all datasets. After standardising the column names and formats, the three data sources were merged into a single consolidated dataframe, forming the initial corpus for further cleaning.

```python
import pandas as pd

# 1. Load all three files
df_nac = pd.read_excel("delfi_nacionaliniai.xlsx")
df_pag = pd.read_excel("delfi_pagrindinis.xlsx")
df_val = pd.read_excel("delfi_valstybiniai.xlsx")


# 2. Rename columns to match structure (URL column differs in each file)
df_nac = df_nac.rename(columns={
    "title": "title",
    "date": "date",
    "text": "text",
    "data-page-selector": "url"   # national exam: URL column
})

df_pag = df_pag.rename(columns={
    "title": "title",
    "date": "date",
    "text": "text",
    "data-page-selector": "url"   # pagrindinis exam: URL column
})

df_val = df_val.rename(columns={
    "title": "title",
    "date": "date",
    "text": "text",
    "data-page-selector": "url"   # valstybiniai exam: URL column
})

 # 3. Keep only necessary columns
 cols = ["title", "date", "text", "url"]

 df_nac = df_nac[cols]
 df_pag = df_pag[cols]
 df_val = df_val[cols]

 # 4. Merge all files
 df = pd.concat([df_nac, df_pag, df_val], ignore_index=True)
```

After that, to ensure data quality and uniformity for reliable sentiment analysis, duplicate articles based on URL were removed, entries with missing or invalid dates were excluded, publication dates were converted into a unified datetime format, and the corpus was filtered to include only texts published between 2020 and 2025. The corpus resulted in 204 articles after cleaning.

```python
# 5. Remove duplicates by URL
df = df.drop_duplicates(subset=["url"])
print("Rows after removing duplicates:", len(df))

# 6. Clean and filter dates (2020–2025)
df["date"] = pd.to_datetime(df["date"], errors="coerce")
df = df.dropna(subset=["date"])

df["year"] = df["date"].dt.year
df = df[df["year"].between(2020, 2025)]

print("Rows after filtering 2020–2025:", len(df))

# 7. Save final merged file
df.to_csv("merged_final.csv", sep=";", index=False, encoding="utf-8")
```

3.3 Lexicon-Based Sentiment Analysis (VADER)

For the sentiment analysis, the cleaned dataset stored in “merged_final.csv” was reloaded:

```
# Load pre-cleaned merged dataset
df = pd.read_csv("merged_final.csv", sep=";")

print("Number of articles:", len(df))
print(df.head(3)[["title", "date", "url"]])

# Parse dates again
df["date"] = pd.to_datetime(df["date"], errors="coerce")
df = df.dropna(subset=["date"])
df["year"] = df["date"].dt.year
```

A lexicon-based approach was then implemented using the VADER (Valence Aware Dictionary and sEntiment Reasoner) model, which produces four scores for each text: negative, neutral, positive, and a combined compound score ranging from −1 to +1. Firstly, the model was initialised and applied to the article texts:

```
sia = SentimentIntensityAnalyzer()

# Apply VADER scores to the article text
scores = df["text"].astype(str).apply(sia.polarity_scores)

df["neg"] = scores.apply(lambda x: x["neg"])
df["neu"] = scores.apply(lambda x: x["neu"])
df["pos"] = scores.apply(lambda x: x["pos"])
df["compound"] = scores.apply(lambda x: x["compound"])
```

Based on the compound score, each article was assigned to one of three sentiment categories using the standard VADER thresholds:

```
def vader_label(c):
    if c >= 0.05:
        return "positive"
    elif c <= -0.05:
        return "negative"
    else:
        return "neutral"

df["sentiment"] = df["compound"].apply(vader_label)
```

The distribution of labels was inspected to obtain an initial overview of emotional tone:

```
print("Sentiment label distribution:")
print(df["sentiment"].value_counts())
print("\nSentiment label percentages (%):")
print((df["sentiment"].value_counts(normalize=True) * 100).round(2))
```

Finally, the enriched dataset was saved for use in subsequent visualisation and analysis as “merged_with_sentiment.csv”.

3.4 Visualisation and Exploratory Analysis

The last stage of the methods pipeline consisted of generating descriptive statistics and visualisations to explore sentiment patterns in the corpus based on the file “merged_with_sentiment.csv”.

The stage included producing several plots using “matplotlib”, namely:

- Overall sentiment distribution (bar chart of counts by label);
- Sentiment distribution by year (stacked bar chart);
- Average compound score by year (line plot);
- Histogram of compound scores (overall polarity distribution);
- Boxplot of compound scores per year (variation and outliers).

Example code for the overall distribution is represented below, while others can be found in the notebook:

```
# 2. Overall sentiment distribution
df["sentiment"].value_counts().plot(kind="bar")
plt.title("Sentiment distribution (VADER)")
plt.xlabel("Sentiment")
plt.ylabel("Number of articles")
plt.tight_layout()
plt.show()
```

Consequently, to gain lexical insight, word clouds were constructed separately for positive and negative subsets using the “wordcloud” library:

```
# 7. WordClouds for positive & negative=
stopwords = set(STOPWORDS)

# Positive articles
pos_text = " ".join(df[df["sentiment"]=="positive"]["text"].astype(str))
if len(pos_text.strip()) > 0:
    wc_pos = WordCloud(width=1100, height=700,
                       background_color="white",
                       stopwords=stopwords).generate(pos_text)
    plt.imshow(wc_pos, interpolation="bilinear")
    plt.axis("off")
    plt.title("WordCloud - Positive Articles")
    plt.tight_layout()
    plt.show()
else:
    print("No positive texts for wordcloud.")

# Negative articles
neg_text = " ".join(df[df["sentiment"]=="negative"]["text"].astype(str))
if len(neg_text.strip()) > 0:
    wc_neg = WordCloud(width=1100, height=700,
                       background_color="white",
                       stopwords=stopwords).generate(neg_text)
    plt.imshow(wc_neg, interpolation="bilinear")
    plt.axis("off")
    plt.title("WordCloud - Negative Articles")
    plt.tight_layout()
    plt.show()
else:
    print("No negative texts for wordcloud.")
```

Finally, the ten most positive and ten most negative articles were identified based on the compound score, to allow qualitative inspection of extreme cases:

```
# 8. Top 10 most positive / negative
top_pos = df.sort_values("compound", ascending=False).head(10)
top_neg = df.sort_values("compound", ascending=True).head(10)
```

Based on the given plots and lists, the main sentiment patterns in Lithuanian examination-related news could be described and interpreted (see Results section).

4. Results

4.1 Overall Sentiment Distribution

Starting with the sentiment distribution across the corpus, it can be seen that the majority of articles were neutral (131; 64.22%), followed by a smaller proportion of negative (69; 33.82%), and very few positive articles (4; 1.96%). This suggests that Delfi.lt generally maintains a balanced tone, with occasional negative sentiment linked to topics further examined in the thematic analysis later in the chapter (Figure 8).

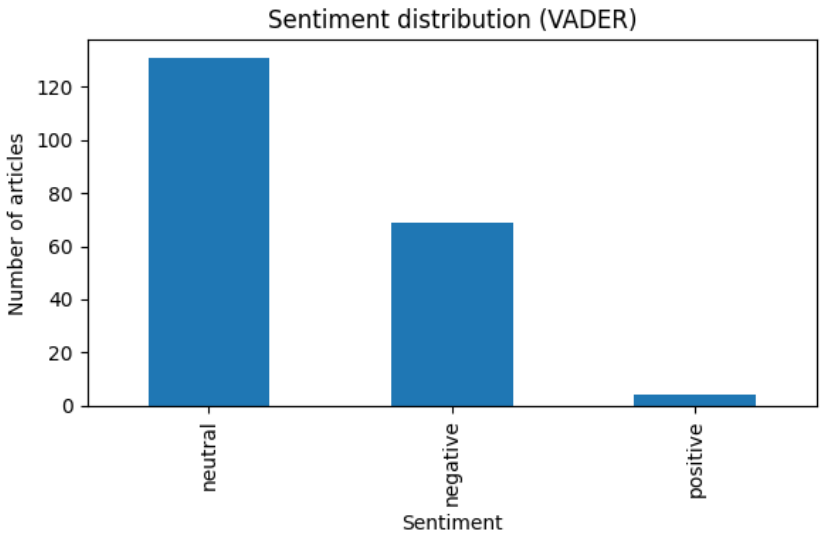

Figure 8. Sentiment distribution of Delfi.lt news articles covering Lithuanian school examinations (2020–2025).

Source: Author's own calculations based on the collected Delfi.lt corpus.

4.2 Temporal Dynamics (2020–2025)

As for temporal dynamics, sentiment varied noticeably across years. In 2020–2022, negative sentiment dominated, accounting for 60% in 2020, 54.5% in 2021, and 58.8% in 2022. These trends may correspond with educational policy reforms and the related public debates, as well as the adaptation to the COVID-19 context. In contrast, 2023 and 2024 showed more balanced patterns, with negative and neutral articles occurring in nearly equal proportions (48% vs. 48% in 2023; 47.1% vs. 52.9% in 2024, in turn). By 2025, the coverage became mostly neutral, with 79.5% of articles classified as neutral and only 19.6% as negative, indicating a shift toward more informational and less emotionally charged reporting (Figure 9).

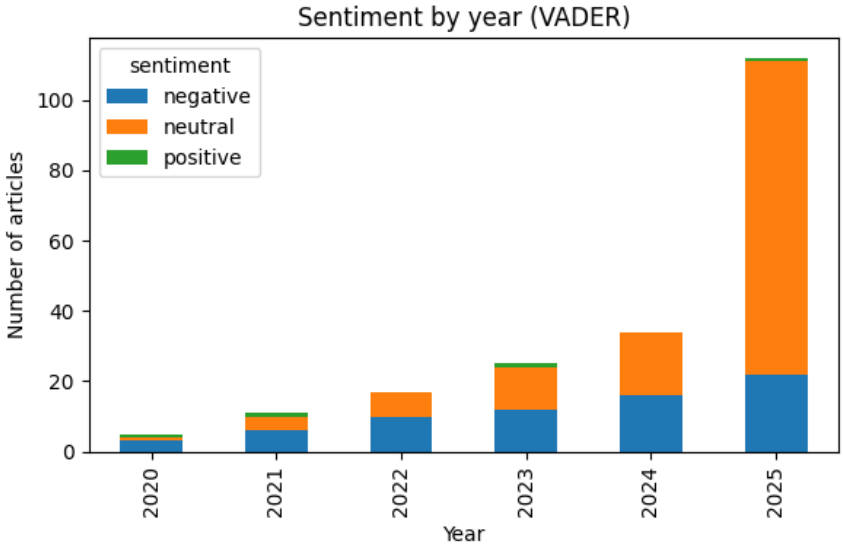

Figure 9. Sentiment distribution by year for Delfi.lt news articles covering Lithuanian school examinations (2020–2025).

Source: Author's own calculations based on the collected Delfi.lt corpus.

4.3 Distribution of Compound Scores

The intensity of emotional polarity was then examined with the histogram of compound scores, which revealed a strong clustering around zero, confirming that most articles were emotionally neutral, with relatively few highly polarised news (Figure 10).

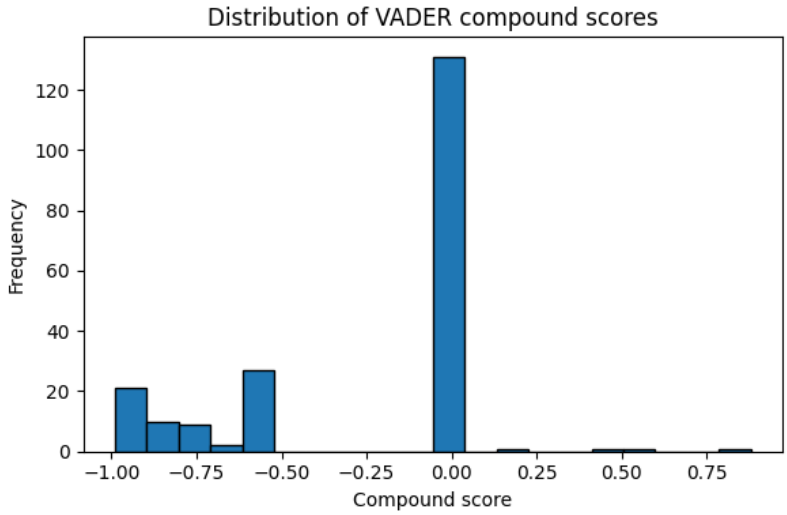

Figure 10. Distribution of VADER compound scores for Delfi.lt news articles on Lithuanian school examinations (2020–2025).

Source: Author's own calculations based on the collected Delfi.lt corpus.

A boxplot of compound scores by year further illustrated variation and highlighted years with greater emotional volatility, particularly 2020–2022 (Figure 11).

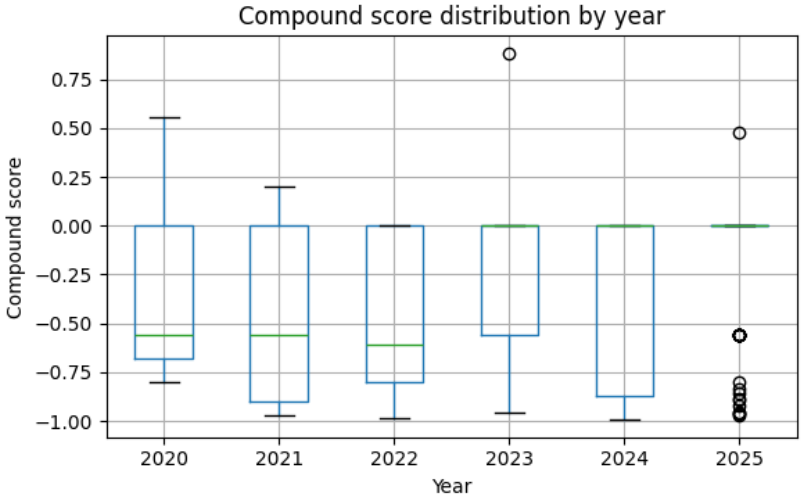

Figure 11. Boxplot of compound sentiment score distribution by year (2020–2025).
Source: Author's own calculations based on the collected Delfi.lt corpus.

4.4 Lexicon-Based Content Insights

Consequently, word clouds for positive and negative subsets revealed clear thematic contrasts, with negative articles frequently highlighting stress, dissatisfaction, exam difficulty, and systemic challenges, while positive narratives centred on student achievements, successful reforms, and ministry initiatives (Figure 12 and 13).

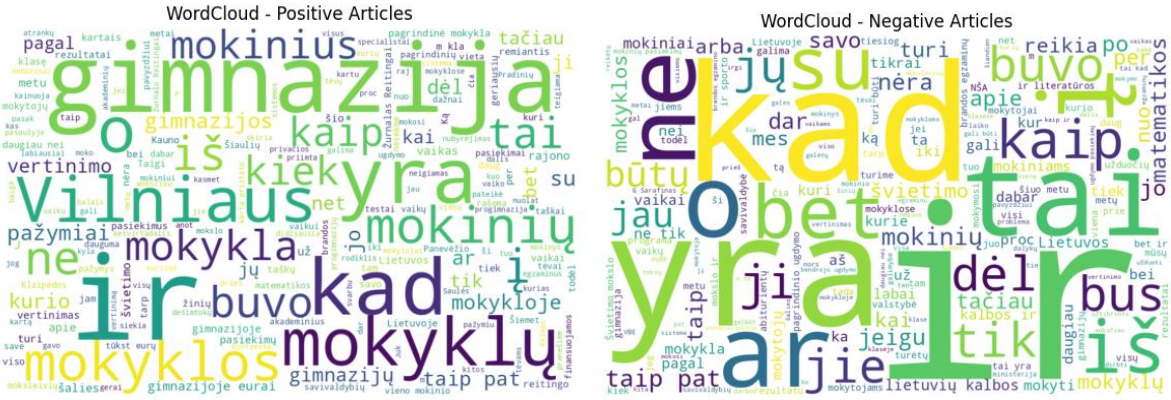

Figure 12 and 13. Word cloud of positive and negative articles.
Source: Author's own compilation based on the collected Delfi.lt corpus.

The top articles in both categories confirmed these trends, with negative articles predominantly reflecting public concern surrounding the Lithuanian examination system, namely systemic instability and "permanent crisis", fears of declining educational standards due to reform outcomes (Appendix 1, Table A1). Further negativity was centred around discussions about directing weaker students into vocational pathways, calls for the resignation of the Minister of Education, and warnings from NŠA about increased student stress due to procedural shifts. As about the positive articles, they were significantly fewer (n = 4) and mainly focused on student achievements and

successful school performance, highlighting exceptional exam results, upward movements in school rankings, and discussions on competencies needed for high-achieving pupils (Appendix 2, Table A2).

5. Conclusions

To conclude, during the sentiment analysis in 204 Lithuanian news articles about national school examinations, the four initial key research questions were addressed:

- First, the analysis showed that the overall sentiment distribution is predominantly neutral, with almost two-thirds of articles classified as neutral and one-third as negative, while positive coverage is almost absent;
- Second, temporal analysis demonstrated that sentiment is not static: negative sentiment peaked during 2020–2022, which may be related to pandemic uncertainty and frequent examination reforms, while later years (2024–2025) returned to predominantly neutral reporting;
- Third, lexical patterns revealed clear thematic contrasts: negative articles emphasised stress, uncertainty, difficulty, and criticism of the examination system, whereas positive articles focused almost exclusively on student achievements and exceptional results (e.g., šimtukai).
- Finally, the evaluation of the lexicon-based approach showed that while VADER provides a functional baseline for Lithuanian texts, it has notable limitations due to the lack of Lithuanian-specific vocabulary, leading to an overclassification of neutrality and reduced sensitivity to nuanced emotional expressions.

Overall, the study successfully answered the research questions and provides an initial computational baseline for analysing Lithuanian educational news. However, the findings highlight the need for improved Lithuanian sentiment resources and more advanced, language-specific models for future work.

6. References

1. Mahander Kumar, Lal Khan and Hsien-Tsung Chang. "Evolving techniques in sentiment analysis: a comprehensive review." PeerJ Computer Science, 11 (2025). https://doi.org/10.7717/peerj-cs.2592.

2. Orestes Appel, F. Chiclana, Jenny Carter and H. Fujita. "A hybrid approach to the sentiment analysis problem at the sentence level." Knowl. Based Syst., 108 (2016): 110-124. https://doi.org/10.1016/j.knosys.2016.05.040.

3. Huang Huang, A. Asemi and Mumtaz Begum Mustafa. "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions." IEEE Access, 11 (2023): 90367-90382. https://doi.org/10.1109/access.2023.3307308.

4. Walaa Medhat, A. Hassan and H. Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal, 5 (2014): 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011.

5. E. Mercha and H. Benbrahim. "Machine learning and deep learning for sentiment analysis across languages: A survey." Neurocomputing, 531 (2023): 195-216. https://doi.org/10.1016/j.neucom.2023.02.015.

6. Alexander Ligthart, C. Catal and B. Tekinerdogan. "Systematic reviews in sentiment analysis: a tertiary study." Artificial Intelligence Review, 54 (2021): 4997 - 5053. https://doi.org/10.1007/s10462-021-09973-3.

7. Lin Yue, Weitong Chen, Xue Li, Wanli Zuo and Minghao Yin. "A survey of sentiment analysis in social media." Knowledge and Information Systems, 60 (2018): 617 - 663. https://doi.org/10.1007/s10115-018-1236-4.

APPENDIX 1

Table A1. Top 10 most negative articles based on VADER compound scores.

№	group	date	title	url	compound	sentiment
1	Negative	12.10.2024	Atviras išskirtinio šalies mokytojo interviu – apie tai, kodėl mokymasis neturi visada būti smagus ir už ką nekenčia „TikTok'ų"	https://www.delfi.lt/news/daily/education/atviras-isskirtinio-salies-mokytojo-interviu-apie-tai-kodel-mokymasis-neturi-visada-buti-smagus-ir-uz-ka-nekencia-tiktok-u-120056280	-0,9898	negative
2	Negative	29.03.2024	Gediminas Merkys. Brandos egzaminų ir tarpinių patikrinimų nuolatinė krizė. Ar ji tikrai neišvengiama?	https://www.delfi.lt/news/ringas/lit/gediminas-merkys-brandos-egzaminu-ir-tarpiniu-patikrinimu-nuolatine-krize-ar-ji-tikrai-neisvengiama-96254609	-0,986	negative
3	Negative	30.12.2022	Svarbiausi 2022 metų Lietuvos įvykiai	https://www.delfi.lt/news/daily/lithuania/svarbiausi-2022-metu-lietuvos-ivykiai-92157511	-0,9831	negative
4	Negative	12.07.2024	Vigilijus Jukna. Švietimo reformomis kelias į pragarą grįstas?	https://www.delfi.lt/news/ringas/politics/vigilijus-jukna-svietimo-reformomis-kelias-i-pragara-gristas-120036033	-0,9762	negative
5	Negative	03.11.2021	Mokinių laukia gausybė pokyčių: planuoja keisti brandos egzaminų tvarką, kai kurių dalykų galbūt visai atsisakys	https://www.delfi.lt/news/daily/lithuania/mokiniu-laukia-gausybe-pokyciu-planuoja-keisti-brandos-egzaminu-tvarka-kai-kuriu-dalyku-galbut-visai-atsisakys-88576285	-0,9705	negative
6	Negative	11.06.2025	Prabilo apie pakeitimus mokyklose – pokyčiai paliestų tūkstančius moksleivių	https://www.delfi.lt/news/daily/education/prabilo-apie-pakeitimus-mokyklose-pokyciai-paliestu-tukstancius-moksleiviu-120117331	-0,9705	negative
7	Negative	29.11.2022	Urbanovič apie PUPP egzaminų pakeitimus: nukreipdami silpniausius moksleivius į profesines mokyklas, grįžtame į sovietmetį	https://www.delfi.lt/news/daily/lithuania/urbanovic-apie-pupp-egzaminu-pakeitimus-nukreipdami-silpniausius-moksleivius-i-profesines-mokyklas-griztame-i-sovietmeti-91879683	-0,9636	negative
8	Negative	15.10.2025	Politologas: švietimo ministrė turėtų atsistatydinti	https://www.delfi.lt/news/daily/politics/politologas-svietimo-ministre-turetu-atsistatydinti-120161360	-0,963	negative
9	Negative	24.08.2023	Švietimo ministerija išvardijo šių mokslo metų naujoves: suplanuotos ir naujos pamokos	https://www.delfi.lt/seima/noriu-i-mokykla/svietimo-ministerija-isvardijo-siu-mokslo-metu-naujoves-suplanuotos-ir-naujos-pamokos-94286853	-0,9584	negative
10	Negative	01.07.2025	NŠA vadovas apie ištesėtus pažadus, pasikeitusią tvarką ir galimą mokinių stresą	https://www.delfi.lt/projektai/studijos/nsa-vadovas-apie-istesetus-pazadus-pasikeitusia-tvarka-ir-galima-mokiniu-stresa-120115630	-0,9584	negative

Source: Author's own calculations based on the collected Delfi.lt corpus.

APPENDIX 2

Table A2. Positive articles based on VADER compound scores (n = 4).

№	group	date	title	url	compound	sentiment
1	Positive	19.01.2023	Specialistė pasakė, ko stokoja vien dešimtukais besimokantys mokiniai: šiuolaikiniame pasaulyje reikia gerokai daugiau	https://www.delfi.lt/seima/noriu-i-mokykla/specialiste-pasake-ko-stokoja-vien-desimtukais-besimokantys-mokiniai-siuolaikiniame-pasaulyje-reikia-gerokai-daugiau-92310299	0,8796	positive
2	Positive	06.05.2020	Mokinių pasiekimų reitingai: prestižinių gimnazijų trejetuke – pokyčiai, kai kurios mokyklos pademonstravo įspūdingą šuolį	https://www.delfi.lt/news/daily/lithuania/mokiniu-pasiekimu-reitingai-prestiziniu-gimnaziju-trejetuke-pokyciai-kai-kurios-mokyklos-pademonstravo-ispudinga-suoli-84202817	0,5574	positive
3	Positive	24.07.2025	Šimtukų fiesta Vilniuje: kas slypi už aukščiausių rezultatų?	https://www.delfi.lt/projektai/studijos/simtuku-fiesta-vilniuje-kas-slypi-uz-auksciausiu-rezultatu-120128602	0,4767	positive
4	Positive	12.05.2021	Gimnazijų reitingas atsakė į klausimą, ar mokymosi lygį lemia savivaldybės gerovė	https://www.delfi.lt/news/daily/lithuania/gimnaziju-reitingas-atsake-i-klausima-ar-mokymosi-lygi-lemia-savivaldybes-gerove-87169965	0,2023	positive

Source: Author's own calculations based on the collected Delfi.lt corpus.